Unsupervised Segmentation and Correspondence

Rehan Ahmad

Department of Physics, LUMS

Contents

1 Introduction

1.1 Background and Relevance

The mathematical study of Stochastic Differential Equations (SDEs) underpins the development of Score-Based Models and Diffusion Models which have gained prominence in generative modeling due to their ability to approximate complex distributions and perform sampling through reverse-time processes. By leveraging the Fokker-Planck equation, Itô calculus, and concepts like Variance Preserving and Variance Exploding SDEs, these models redefine generative tasks by connecting probabilistic distributions with stochastic processes.Understanding these foundations not only elucidates how data perturbations and score estimations work but also highlights the deep connections between physics-inspired models and modern machine learning.

2 Introduction to Stochastic Differential Equations

2.1 Setting up the framework

Imagine you are modeling the growth of a population $x(t)$ of bacteria. In a deterministic setting, we might use the following differential equation to model the growth of the population:

$$\frac{dx}{dt} = rx \tag{1}$$

where r is the growth rate. This equation tells us that the rate of change of the population is proportional to its current size. Given an initial population $x(0)$, the future population at any time t is completely determined. The solution to this equation is trivial:

$$x(t) = x(0)e^{rt}. \tag{2}$$

Now, suppose the growth rate r is not constant but fluctuates due to random environmental factors such as changes in nutrient availability or temperature. In this case, the change in x is no longer deterministic. This is the subject of stochastic differential equations (SDEs).

To get a flavor of SDEs, let us begin with a linear ordinary differential equation with a deterministic driving force that describes a damped harmonic oscillator under an external force $f(t)$:

$$\frac{dx}{dt} = -\gamma x + f(t). \tag{3}$$

We can discretize this ODE and write it in terms of differentials:

$$\Delta x(t_n) = x(t_n)\Delta t + f(t_n)\Delta t \tag{4}$$

where $f(t_n)\Delta t$ is the driving term. The value of $x(t_n + \Delta t)$ is then given by:

$$x(t_n + \Delta t) = x(t_n) + \Delta x(t_n)$$

Substituting $\Delta x(t_n)$ from (4), we have:

$$x(t_n + \Delta t) = x(t_n) + x(t_n)\Delta t + f(t_n)\Delta t. \tag{5}$$

If we know the value of x at $t = 0$, then:

$$x(\Delta t) = x(0)(1 + \Delta t) + f(0)\Delta t. \tag{6}$$

What we are particularly interested in is the scenario where the driving term $f(t_n)\Delta t$ becomes random at each time step t_n. This involves replacing $f(t_n)$ with a random variable y_n at each t_n. As a result, the difference equation (4) transforms into the following:

$$\Delta x(t_n) = x(t_n)\Delta t + y_n \Delta t \tag{7}$$

where $y_n \sim P(Y = y_1, \ldots, y_i)$.

This is called a *stochastic difference equation.* It states that at each time t_n, we pick a value for the random variable y_n sampled from its probability density $P(Y)$ and add $y_n\Delta t$ to $x(t_n)$. Consequently, we can no longer predict the exact value of x at some future time T in advance. Instead, $x(T)$ depends on the cumulative effect of all random increments y_n up to time T, which are determined sequentially as the process evolves.

The solution for x at time Δt is:

$$x(\Delta t) = x(0)(1 + \Delta t) + y_0 \Delta t \tag{8}$$

Thus, $x(\Delta t)$ is now a random variable. If the initial condition $x(0)$ is fixed (i.e., not random), then $x(t)$ can be expressed as a linear transformation of the random variable y_0, which represents the noise increment over the interval $[0, \Delta t]$. On the other hand, if $x(0)$ is also a random variable, then $x(t)$ is a linear combination of two random variables: the initial condition $x(0)$ and the noise increment y_0.

When we proceed to the next time step to calculate $x(2\Delta t)$, this value depends on $x(0)$, y_0, and the noise increment for the second step y_1. At each time step, the solution of the stochastic difference equation $x(t_n)$ is a random variable, and this random variable evolves as time progresses. Thus, solving a stochastic difference equation requires determining the probability density of $x(t_n)$ for all future times t_n. This process involves deriving the probability density of $x(t_n)$ from:

- The probability densities of the noise increments y_n.
- The probability density of $x(0)$, if $x(0)$ is random.

From Stochastic Difference Equations to Stochastic Differential Equations

Stochastic differential equations (SDEs) are obtained by taking the limit $\Delta t \to 0$ in stochastic difference equations. In this continuous-time framework, the solution to an SDE is characterized not by a single trajectory but by a probability density function that describes the value of x at future times t. Just as with ordinary (deterministic) differential equations, finding a closed-form solution to an SDE is not always possible. However, in many simple cases, explicit solutions can be derived. For more complex systems, numerical methods or approximations are typically employed to study the evolution of the probability density over time.

In addition to obtaining the probability density for x at future times t_n, we can also ask how x evolves with time given a specific set of values for the random increments y_n.

This focuses on two key aspects:

- The probability density of x at specific times.
- The sample paths of x, which describe specific realizations of its evolution under different noise realizations.

Realization of the Noise : Let y_n represent the random noise increments (e.g., increments of a Wiener process) over discrete time intervals $[\Delta t, 2\Delta t, \dots]$. A realization of the noise is a specific set of sampled values $\{y_1, y_2, \dots, y_n\}$ drawn from the noise's probability density $y_n \sim N(0, \Delta t)$

Sample Path of $x(t)$: A sample path is a trajectory of $x(t)$ over time that corresponds to a specific realization of the noise. The full solution to the SDE is the collection of all possible sample paths, along with their associated probabilities. Mathematically, this corresponds to a probability measure over the space of functions $x(t)$. This full solution is rarely needed in practice. Instead, we focus on:

- The marginal probability density $P(x,t)$: The probability density of $x(t)$ at each time t, obtained from the Fokker-Planck equation associated with the SDE.
- Correlation Functions: Quantities like $\mathbb{E}[x(t)x(t')]$, which describe how $x(t)$ at one time is statistically related to $x(t')$ at another.

2.2 Introduction to Weiner Increments

By "Gaussian noise," we mean that each of the random increments $y_0\Delta t$ has a Gaussian probability density. Specifically, $y_n \sim \mathcal{N}(0, \sigma^2)$. First, consider the simplest stochastic difference equation where the increment of x consists solely of the random increment $y_n\Delta t$. This reduces equation (7) to:

$$\Delta x(t_n) = y_n \Delta t \tag{9}$$

In literature, Gaussian noise is often referred to as *Wiener noise*, and the random increment is expressed as:

$$\Delta W_n = y_n \Delta t \tag{10}$$

The discrete differential equation for x can thus be written as:

$$\Delta x(t_n) = \Delta W_n \tag{11}$$

Each Wiener increment ΔW_n is independent of the others and has the same probability density:

$$P(\Delta W) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\Delta W)^2}{2\sigma^2}} \tag{12}$$

This density is Gaussian with zero mean, and we set the variance as $\sigma^2 = V = \Delta t$. This choice is critical and we will explain the reason for it in the next section. For simplicity, we often denote a Wiener increment in a given time step Δt as ΔW, without referencing the subscript n, since all increments share the same distribution and are independent.

Solving the Difference Equation for x

We can solve the difference equation for x by starting with $x(0) = 0$ and repeatedly adding Δx. The solution is:

$$x_n \equiv x(n\Delta t) = \sum_{i=1}^{n-1} \Delta W_i \tag{13}$$

The probability density of x_n can now be calculated. Since the sum of Gaussian random variables is also Gaussian, the probability density of x_n is Gaussian (a property that is proved in most introductory books). The mean and variance of x_n are the sums of the means and variances of ΔW_i, respectively, because the ΔW_i are independent. Thus:

$$\mu = \langle x_n \rangle = 0 \tag{14}$$

$$V(x_n) = n\sigma^2 = n\Delta t \tag{15}$$

The solution to the difference equation is then:

$$P(x_n) = \frac{1}{\sqrt{2\pi V}} e^{-\frac{x_n^2}{2V}} = \frac{1}{\sqrt{2\pi n \Delta t}} e^{-\frac{x_n^2}{2n\Delta t}} \tag{16}$$

From Difference Equation to Differential Equation

Transforming the difference equation (11) into a differential equation gives the simplest SDE:

$$dx = dW \tag{17}$$

To solve this, consider T future steps with N discrete time steps and take the limit as $N \to \infty$:

$$x(T) = \lim_{N\to\infty} \sum_{i=0}^{N-1} \Delta W_i = \int_0^T dW(t) = W(T) \tag{18}$$

Here, the stochastic integral $W(T) = \int_0^T dW(t)$ is defined as the limit of the sum of all increments of the Wiener process.

Stochastic Integral: A stochastic integral, as the sum of random variables, is itself a random variable. In many cases, we can compute its probability density.

For $x(T)$, the probability density is Gaussian because it sums independent Gaussian variables. The mean is zero since each variable has zero mean. The variance of $x(T)$

does not require the limit $N \to \infty$ because N factors out:

$$V(x(T)) = \sum_{i=1}^{N-1} V[\Delta W_i] = \sum_{i=1}^{N-1} \Delta t = N\Delta t = N(T/N) = T \tag{19}$$

Thus, the probability density of $W(T)$ is:

$$P(W(T)) = P(x(T)) = P(x,T) = \frac{1}{\sqrt{2\pi T}} e^{-\frac{x^2}{2T}} \tag{20}$$

Wiener Process The Wiener process is defined as $W(T)$, while dW represents an increment of the Wiener process:

$$W(T) = \int_0^T dW \tag{21}$$

Note: While we often refer to dW as the Wiener process, the Wiener process is technically $W(T)$, and dW is its increment.

2.3 The variance of a wiener increment must satisfy Δt

So far, when considering Wiener increments, we have assumed that the variance of Wiener increments is Δt, which led to the fact that the variance of $x(T)$ is proportional to T. It turns out that setting the variance to any other value leads to an unphysical system. To show this, we set $V[\Delta W(\Delta t)] = \Delta t^\alpha$ and calculate the variance of $x(T)$ once again:

$$V(x(T)) = \sum_{i=1}^{N-1} V[\Delta W_i] = N(\Delta t)^\alpha = N\left(\frac{T}{N}\right)^\alpha = N^{(1-\alpha)}T^\alpha \tag{22}$$

Now we take the continuum limit $N \to \infty$ to obtain a stochastic differential equation. When $\alpha > 1$, we have:

$$\lim_{N\to\infty} V(x(T)) = T^\alpha \lim_{N\to\infty} N^{1-\alpha} = 0 \tag{23}$$

And when $\alpha < 1$, we have:

$$\lim_{N\to\infty} V(x(T)) = T^\alpha \lim_{N\to\infty} N^{1-\alpha} = \infty \tag{24}$$

Neither of these results make sense for obtaining a stochastic differential equation that describes real systems driven by noise. Thus, we are forced to choose $\alpha = 1$ and hence $V[\Delta W(\Delta t)] \propto \Delta t$.

2.4 General SDE Form and an Example of SDE

A general SDE for a single variable X_t can be written as:

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t \tag{25}$$

Where:

- X_t is the stochastic process.
- $\mu(X_t, t)$ is the drift term, representing the deterministic part of the change.
- $\sigma(X_t, t)$ is the diffusion term, representing the stochastic part of the change.
- dW_t is the increment of the Wiener process.

Since the variance of dW_t must be proportional to dt, and any constant of proportionality can always be absorbed into $\sigma(X_t, t)$, the variance of dW_t is defined to be equal to dt. Therefore, the probability density of dW_t is:

$$P(dW) = \frac{1}{\sqrt{2\pi dt}} e^{-\frac{(dW)^2}{2dt}} \tag{26}$$

Building on the above, let us consider a simple SDE of the form:

$$dX = e^t dW \tag{27}$$

Where $W(t)$ is a Wiener process. This SDE can be expressed as:

$$dX = f(t)dW \tag{28}$$

To solve this equation, we integrate both sides from 0 to t:

$$X(t) - X(0) = \int_0^t e^s dW(s) \tag{29}$$

Assuming the initial condition $X(0) = 0$, we have:

$$X(t) = \int_0^t e^s dW(s) \tag{30}$$

This integral represents the accumulation of the stochastic process e^s weighted by the increments of the Wiener process over $[0, t]$.

Mean of $X(t)$: To understand the properties of $X(t)$, let us first calculate its mean. Since the integrand e^s and the Wiener process increments $dW(s)$ are independent and the expectation of $dW(s)$ is zero:

$$\mathbb{E}[X(t)] = \mathbb{E}\left[\int_0^t e^s dW(s)\right] = 0 \tag{31}$$

Variance of $X(t)$: The variance of $X(t)$ can be calculated using Itô isometry:

$$V(X(t)) = \mathbb{E}\left[\left(\int_0^t e^s dW(s)\right)^2\right] = \int_0^t (e^s)^2 ds = \int_0^t e^{2s} ds \tag{32}$$

Evaluating the integral, we obtain:

$$V(X(t)) = \frac{e^{2t} - 1}{2} \tag{33}$$

3 Ito Calculus

Consider the solution to the simple differential equation:

$$dx = -\theta x \, dt \tag{34}$$

In elementary courses on Ordinary Differential Equations (ODEs), the solution to this equation is straightforward. The deterministic nature of this equation ensures a well-defined trajectory for $x(t)$. However, when we introduce stochastic elements, the rules for solving such equations become less straightforward, as we will see.

To understand this, let us first revisit the solution to the deterministic equation (34) . The value of x at time $t + dt$ is the value at time t plus the infinitesimal change dx:

$$x(t + dt) = x(t) - \theta x(t)dt = (1 - \theta dt)x(t) \tag{35}$$

From the Taylor series, $e^{\theta dt} \approx 1 + \theta dt$. We can therefore write the equation for $x(t + dt)$ as:

$$x(t + dt) = e^{-\theta dt}x(t) \tag{36}$$

This tells us that to move x from time t to $t + dt$, we merely have to multiply $x(t)$ by the factor $e^{-\theta dt}$. So to move by two lots of dt, we simply multiply by this factor twice:

$$x(t + 2dt) = e^{-\theta dt}x(t + dt) = e^{-\theta dt}[e^{-\theta dt}x(t)] = e^{-\theta \cdot 2dt}x(t). \tag{37}$$

To obtain $x(t + \tau)$, all we have to do is apply this relation repeatedly. Let us say that $dt = \tau/N$ for N as large as we want. Thus, dt is a small but finite time-step, and we can make it as small as we want. That means that to evolve x from time t to $t + \tau$, we can apply (36) N times:

$$x(t+\tau) = \left(e^{-\theta dt}\right)^N x(t) = e^{-\theta \sum_{n=1}^{N} dt} x(t) = e^{-\theta N dt} x(t) = e^{-\theta \tau} x(t) \tag{38}$$

is the solution to the differential equation. If γ is a function of time, so that the equation becomes:

$$dx = -\theta(t) x\, dt \tag{39}$$

As before, we set $dt = \tau/N$ so that it is a small finite-time step, but this time we have to explicitly take the limit as $N \to \infty$ to obtain the solution to the differential equation:

$$\begin{aligned} x(t+\tau) &= \lim_{N\to\infty} \prod_{n=1}^{N} \left(e^{-\theta(t+ndt)dt}\right) x(t) \\ &= \lim_{N\to\infty} e^{-\sum_{n=1}^{N} \theta(t+ndt)dt} x(t) \\ &= e^{-\int_t^{t+\tau} \theta(t)dt} x(t) \end{aligned} \tag{40}$$

Notice that we were able to solve the equation (36) using $e^{\alpha dt} \approx 1 + \alpha dt$. The approximation $e^{\alpha dt} \approx 1 + \alpha dt$ works because the terms in the power series expansion for $e^{\alpha dt}$ that are second-order or higher in dt ($dt^2, dt^3, \ldots$) will vanish in comparison to dt as $dt \to 0$. The result of being able to ignore terms that are second-order and higher in the infinitesimal increment leads to the usual rules for differential equations. (It also means that any equation we write in terms of differentials dx and dt can alternatively be written in terms of derivatives.)

However, we will now spend some time showing that as $dt \to 0$, $(dW)^2$ does not tend to 0. We must therefore learn a new rule for the manipulation of stochastic differential equations.

To examine whether $(dW)^2$ makes a non-zero contribution to the solution, we sum $(dW)^2$ over all the time-steps for a finite time T. To do this, we will return to a discrete description so that we can explicitly write down the sum and then take the continuum limit. That is, we examine the limit:

$$\Delta X = \lim_{N\to\infty} \sum_{i=1}^{N} (dW_i)^2 \tag{41}$$

The first thing to note is that the expectation value of $(\Delta W)^2$ is equal to the variance of ΔW, because $\langle \Delta W \rangle = 0$. From $\mathrm{Var}(\Delta W) = \langle (\Delta W)^2 \rangle - \langle \Delta W \rangle^2$, we have:

$$\langle \Delta W^2 \rangle = \Delta t \tag{42}$$

This tells us immediately that the expectation value of $(\Delta W)^2$ does not vanish with respect to the time-step Δt, and so the sum of these increments will not vanish when we sum over all the time-steps and take the infinitesimal limit (because we would have a Gaussian on our hands with a variance Δt). In fact, the expectation value of the sum of all the increments $(dW)^2$ from 0 to T is simply T:

$$\langle\int_0^T (dW)^2\rangle = \int_0^T \langle(dW)^2\rangle = \int_0^T dt = T \tag{43}$$

The Integral of $(\Delta W)^2$ is Deterministic:

In order to prove that the integral of $(\Delta W)^2$ is deterministic, we will look at the variance $\int_0^T (dW)^2$ and show that it is zero. To do this, we will first argue about the differential $(\Delta W)^2$ itself. We found that the expectation of $(\Delta W)^2$ is proportional to Δt, and so we have a strong hunch that the variance of $(\Delta W)^2$ must be proportional to $(\Delta t)^2$. Think in terms of *Hint*: $\mathrm{Var}(x) = \mathbb{E}[x^2] - (\mathbb{E}[x])^2$. Let us now explicitly calculate the variance of $(\Delta W)^2$ using the probability density for ΔW:

$$P(\Delta W) = \frac{1}{\sqrt{2\pi\Delta t}} e^{-\frac{(\Delta W)^2}{2\Delta t}} \tag{44}$$

Using the change of variable theorem, we now find the probability distribution for $(\Delta W)^2$. We define $z = (\Delta W)^2$. For this, we have two cases: $\Delta W = \sqrt{z}$ and $\Delta W = -\sqrt{z}$. The absolute value of the derivative $\frac{d\Delta W}{dz}$ for $z = (\Delta W)^2$ is:

$$\left|\frac{d\Delta W}{dz}\right| = \left|\frac{d}{dz}(\pm\sqrt{z})\right| = \frac{1}{2\sqrt{z}}.$$

Using the change of variables formula for PDFs, we have:

$$f_z(z) = f_{\Delta W}(\sqrt{z})\left|\frac{d\Delta W}{dz}\right| + f_{\Delta W}(-\sqrt{z})\left|\frac{d\Delta W}{dz}\right|.$$

Since $f_{\Delta W}(x)$ is even, $f_{\Delta W}(-\Delta W) = f_{\Delta W}(\Delta W)$. Thus:

$$f_z(z) = \frac{f_{\Delta W}(\sqrt{z})}{\sqrt{z}}.$$

Substituting the original PDF, we obtain:

$$f_z((\Delta W)^2) = \frac{1}{\sqrt{2\pi\Delta t z}} e^{-\frac{z}{2\Delta t}}. \tag{45}$$

This is a chi-squared distribution with one degree of freedom, scaled by a factor of $2\Delta t$. The variance of the distribution is:

$$V((\Delta W)^2) = 2(\Delta t)^2 \tag{46}$$

The variance of the sum of all the $(\Delta W)^2$ is:

$$V\left[\sum_{n=1}^{N-1}(\Delta W)^2\right] = \sum_{n=1}^{N-1} V((\Delta W)^2) = \sum_{n=1}^{N-1} 2(\Delta t)^2 = 2N\left(\frac{T}{N}\right)^2 = \frac{2T^2}{N} \tag{47}$$

Finally, we can perform what we set out to do, namely evaluating the integral in equation (41):

$$\lim_{N\to\infty}\left[\sum_{n=1}^{N-1} V((\Delta W)^2)\right] = \lim_{N\to\infty}\frac{2T^2}{N} = 0. \tag{48}$$

Since the integral of all the $(dW)^2$ is deterministic, it is equal to its mean T. That is:

$$\int_0^T (dW)^2 = T = \int_0^T dt \tag{49}$$

Thus, we have the surprising result:

$$dW^2 = dt \tag{50}$$

This is officially known as *Ito's rule*. It is the fundamental rule for solving stochastic differential equations that contain Gaussian noise.

3.1 Change of Variables in an SDE

Due to the Ito rule established in equation (50), we will have to keep all terms that are first order in dt and dW, as well as all terms that are second order in dW. In fact, wherever we find terms that are second order in dW, we can simply replace them with dt. To see how this works, consider a simple example in which we want to know the differential equation for:

$$y = x^2 \tag{51}$$

In that case:

$$\begin{aligned} dy &= y(t+dt) - y(t) \\ &= x(t+dt)^2 - x(t)^2 \\ &= (x+dx)^2 - x^2 \\ &= x^2 + 2xdx + (dx)^2 - x^2 \\ &= 2xdx + (dx)^2 \end{aligned} \tag{52}$$

Had x been deterministic, $(dx)^2$ would vanish in the continuum limit, and we would have the usual rule of calculus:

$$dy = 2xdx \quad \text{or} \quad \frac{dy}{dx} = 2x \tag{53}$$

However, if X is a random variable obeying the stochastic differential equation:

$$dX = f dt + g dW \tag{54}$$

then the equation (50) becomes the following:

$$\begin{aligned} dY &= 2x dX + (dX)^2 \\ &= 2x(f dt + g dW) + (f dt + g dW)^2 \end{aligned} \tag{55}$$

Expanding the square term, we have:

$$(f dt + g dW)^2 = f^2 dt^2 + g^2 dW^2 + 2fg dt dW.$$

Just like in normal calculus, we can ignore cross-differentials like $dt\, dW$ and dt^2.

$$(f dt + g dW)^2 \approx g^2 dW^2.$$

We then have:

$$dY = 2fX dt + 2xg dW + g^2 dW^2. \tag{56}$$

Using Ito's rule in equation (50):

$$dY = (2fX + g^2)dt + 2xg dW \tag{57}$$

This is Ito's rule in action.

Ito's Lemma: Fortunately, there is a simple way to calculate the increment of any nonlinear function $y(X)$ in terms of the first and second powers of the increment of X, where X is a random variable. All we have to do is use the Taylor series expansion for $y(X)$, truncated at the second term:

$$dy(X) = \frac{dy}{dX} dX + \frac{1}{2} \frac{d^2 y}{dX^2} (dX)^2 \tag{19}$$

If y is also an explicit function of time as well as X, then this becomes:

$$dy(t, X) = \frac{\partial y}{\partial X} dX + \frac{\partial y}{\partial t} dt + \frac{1}{2} \frac{\partial^2 y}{\partial X^2} (dX)^2 \tag{20}$$

4 The Ornstein-Uhlenbeck Process

The Ornstein-Uhlenbeck (OU) process is a type of stochastic process used to model mean-reverting behavior. It was introduced by Leonard Ornstein and George Eugene Uhlenbeck in 1930 to describe the velocity of a particle undergoing Brownian motion under the influence of friction.

4.1 Understanding Mean-Reverting Behavior

Imagine a particle diffusing in a liquid where there is some friction or restoring force that pulls the particle back towards a central point (mean position) rather than allowing it to drift indefinitely. In this context, the position X_t of the particle at time t can be described by a stochastic differential equation governed by two terms, $\theta(\mu - X_t)$ and σdW_t:

$$dX_t = \theta(\mu - X_t)dt + \sigma dW_t \tag{21}$$

Here:

- θ is the strength of the restoring force, causing the particle to revert to the mean position μ,
- σ is the volatility and represents the intensity of the random fluctuations.

To understand how mean-reverting dynamics would play out, suppose the particle starts at the position X_t far from the mean μ. For example, if μ is at the origin (0) and the particle is initially at $X_0 = 5$, the term $\theta(\mu - X_t)dt$ acts as a restoring force that pulls the particle back towards the mean position μ. The larger θ is, the stronger the pull back to μ. Despite the mean-reverting force, the particle is subject to random fluctuations due to the noise term σdW_t. This represents the random kicks from the surrounding fluid molecules.

Visually, processes obeying (21) look like the following:

[Include plot illustrating sample paths of the Ornstein-Uhlenbeck process]

Here we can see that even though two of the three sample paths start far away from the mean, they quickly converge to a region around the mean. Once in the vicinity of the mean, they move about it in arcs through the momentum factor.

4.2 Rewriting the Ornstein-Uhlenbeck Equation

First, we define a new random variable Y_t:

$$Y_t = X_t - \mu \tag{22}$$

The differential element of dY_t is given by:

$$\begin{aligned} dY_t &= dX_t && (58) \\ &= \theta(\mu - X_t)dt + \sigma dW_t && (59) \\ &= -\theta(X_t - \mu)dt + \sigma dW_t && (60) \\ &= -\theta Y_t dt + \sigma dW_t && (23) \end{aligned}$$

Equation (23) is the reason why we say that the Ornstein-Uhlenbeck equation is governed by "additive noise." The term "additive noise" refers to the fact that the noise does not itself depend on Y_t, but is merely added to the other terms in the equation for dx.

This equation is called the Ornstein-Uhlenbeck equation, and its solution is called the Ornstein-Uhlenbeck process.

4.3 Solution of the Ornstein-Uhlenbeck Process

The next step is to simplify the equation by defining another random variable Z_t as a function of Y_t:

$$Z_t = e^{\theta t} Y_t \tag{24}$$

Using Ito's Lemma (20), we have:

$$dZ_t = \frac{\partial Z_t}{\partial Y_t} dY_t + \frac{\partial Z_t}{\partial t} dt + \frac{1}{2} \frac{\partial^2 Z_t}{\partial Y_t^2} (dY_t)^2 \tag{61}$$

Since $\frac{\partial^2 Z_t}{\partial Y_t^2} = 0$, this reduces to:

$$dZ_t = \frac{\partial Z_t}{\partial Y_t} dY_t + \frac{\partial Z_t}{\partial t} dt \tag{25}$$

Let us apply this:

$$dZ_t = \theta e^{\theta t} Y_t dt + e^{\theta t} dY_t \tag{62}$$

$$= \theta e^{\theta t} Y_t dt + e^{\theta t} (-\theta Y_t dt + \sigma dW_t) \tag{63}$$

$$= e^{\theta t} \sigma dW_t \tag{26}$$

Equation (26) is easy to solve. To do so, we merely sum all the stochastic increments dW over a finite time t, noting that each one is multiplied by $e^{\theta t}\sigma$. Thus, the integral form is:

$$Z_t = Z_S + \int_S^T e^{\theta t} \sigma dW_t \tag{27}$$

Here, S is the start of the integration through time. Now that we have a solution for Z_t, we substitute back to find X_t.

4.4 Final Solution

Substituting $Y_t = e^{-\theta t} Z_t$ back, we have:

$$
\begin{aligned}
Y_t &= e^{-\theta t} Z_t && (64)\\
&= e^{-\theta t}\left(Z_S + \int_S^T e^{\theta t}\sigma dW_t\right) && (65)\\
&= e^{-\theta t} e^{\theta S} Y_S + \sigma \int_S^T e^{\theta(t-T)} dW_t && (28)
\end{aligned}
$$

Substituting $Y_t = X_t - \mu$:

$$X_T = \mu + e^{-\theta(T-S)}(X_S - \mu) + \sigma \int_S^T e^{\theta(t-T)} dW_t \tag{29}$$

Starting from $S = 0$, we obtain:

$$X_T = \mu + e^{-\theta T}(X_0 - \mu) + \sigma \int_0^T e^{-\theta(T-s)} dW_s \tag{30}$$

4.5 Mean and Variance of X_T

To completely determine X_T in equation (30), note that the stochastic integral represents the sum of Gaussian random variables. Thus, all we need to do is calculate its mean and variance.

Mean of X_T:

$$
\begin{aligned}
\mathbb{E}[X_T] &= \mathbb{E}\left[\mu + e^{-\theta T}(X_0 - \mu) + \sigma \int_0^T e^{-\theta(T-s)} dW_s\right] && (66)\\
&= \mu + e^{-\theta T}(X_0 - \mu) + \mathbb{E}\left[\sigma \int_0^T e^{-\theta(T-s)} dW_s\right] && (67)
\end{aligned}
$$

Since $\int_0^T e^{-\theta(T-s)} dW_s$ is a Wiener process and has an expectation of 0:

$$\mathbb{E}[X_T] = \mu + e^{-\theta T}(X_0 - \mu) \tag{31}$$

Variance of X_T: The variance of X_T is:

$$\mathrm{Var}(X_T) = \mathbb{E}[(X_T - \mathbb{E}[X_T])^2] \tag{68}$$

Substituting $\mathbb{E}[X_T]$ from equation (31) and expanding:

$$\mathrm{Var}(X_T) = \mathbb{E}\left[\left(\sigma \int_0^T e^{-\theta(T-s)} dW_s\right)^2\right] \tag{69}$$

$$= \sigma^2 \int_0^T e^{-2\theta(T-s)} ds \quad \text{(by Ito isometry)}. \tag{70}$$

Evaluating the integral:

$$\mathrm{Var}(X_T) = \sigma^2 \int_0^T e^{-2\theta(T-s)} ds \tag{71}$$

$$= \sigma^2 \left[\frac{1}{2\theta} e^{-2\theta(T-s)}\right]_0^T \tag{72}$$

$$= \frac{\sigma^2}{2\theta}\left(1 - e^{-2\theta T}\right) \tag{32}$$

Stationary Distribution: Applying the limit $T \to \infty$, we recover the stationary distribution:

$$\lim_{T\to\infty} \mathbb{E}[X_T] = \mu \tag{33}$$

$$\lim_{T\to\infty} \mathrm{Var}(X_T) = \frac{\sigma^2}{2\theta} \tag{34}$$

5 The Full Linear Stochastic Equation

The general stochastic linear equation reads as:

$$dS_t = -\mu_t S_t\, dt + \sigma_t S_t\, dW_t \tag{35}$$

Here, S_t depends on W_t and does not have explicit time dependence. This equation is called *Geometric Brownian Motion*. We can rewrite the above equation as:

$$\frac{dS_t}{S_t} = \mu_t\, dt + \sigma_t\, dW_t \tag{36}$$

Notice that as S_t approaches zero, so does the change (else the LHS blows up). This effectively limits S_t to positive values, $S_t > 0$.

5.1 Log-Derivative Trick

The question is, as so often with differential equations, whether there exists an analytic solution. To show this, we will examine the quantity $\frac{dS_t}{S_t}$ and apply the stochastic version of the log-derivative trick. The quantity $\frac{dS_t}{S_t}$ has striking similarity to:

$$\frac{\partial \ln S(x)}{\partial x} = \frac{1}{S(x)}\frac{\partial S(x)}{\partial x} \tag{37}$$

However, since we are working with stochastic processes, we cannot apply regular calculus but must use Ito's Lemma:

$$d\ln S_t = \frac{\partial \ln S_t}{\partial t}\,dt + \frac{\partial \ln S_t}{\partial S_t}\,dS_t + \frac{1}{2}\frac{\partial^2 \ln S_t}{\partial S_t^2}(dS_t)^2 \tag{38}$$

Here:

$$\frac{\partial S_t}{\partial t} = 0,$$
$$\frac{\partial \ln S_t}{\partial S_t} = \frac{1}{S_t},$$
$$\frac{\partial^2 \ln S_t}{\partial S_t^2} = -\frac{1}{S_t^2}.$$

Substituting these derivatives into equation (38), we get:

$$d\ln S_t = \frac{1}{S_t}\,dS_t - \frac{1}{2}\frac{1}{S_t^2}(dS_t)^2 \tag{73}$$

5.2 Expanding the Square Term

Plugging equation (35) into dS_t, we expand the square term:

$$(-\mu_t S_t\,dt + \sigma_t S_t\,dW_t)^2 = \mu_t^2 S_t^2\,dt^2 + \sigma_t^2 S_t^2\,dW_t^2 - 2\mu_t\sigma_t S_t^2\,dt\,dW_t.$$

Ignoring terms like dt^2 and $dt\,dW_t$, we approximate:

$$(-\mu_t S_t\,dt + \sigma_t S_t\,dW_t)^2 \approx \sigma_t^2 S_t^2\,dW_t^2. \tag{74}$$

Using Ito's rule ($dW^2 = dt$), this simplifies to:

$$(-\mu_t S_t\,dt + \sigma_t S_t\,dW_t)^2 \approx \sigma_t^2 S_t^2\,dt \tag{39}$$

5.3 Logarithmic Form

Plugging this back into $d\ln S_t$:

$$d\ln S_t = \frac{1}{S_t}\,dS_t - \frac{1}{2}\sigma_t^2\,dt. \tag{75}$$

Rewriting:

$$\frac{1}{S_t}\,dS_t = d\ln S_t + \frac{1}{2}\sigma_t^2\,dt \tag{40}$$

Plugging this into equation (36):

$$\mu_t\,dt + \sigma_t\,dW_t = d\ln S_t + \frac{1}{2}\sigma_t^2\,dt. \tag{76}$$

Integrating both sides:

$$\ln S_t - \ln S_0 = \int_0^t \mu_s\,ds - \frac{1}{2}\int_0^t \sigma_s^2\,ds + \int_0^t \sigma_s\,dW_s. \tag{77}$$

Simplifying:

$$S_t = S_0 \exp\left(\int_0^t \mu_s\,ds - \frac{1}{2}\int_0^t \sigma_s^2\,ds + \int_0^t \sigma_s\,dW_s\right) \tag{41}$$

6 The Stationary Auto-Correlation Function $g(\tau)$

For a stochastic process, it is often useful to know how correlated the values of the process are at two different times. This will tell us how long it takes the process to forget the value it had at some earlier time. We are therefore interested in calculating the correlation coefficient:

$$C_{X(t)X(t+\tau)} = \frac{\langle X(t)X(t+\tau)\rangle - \langle X(t)\rangle\langle X(t+\tau)\rangle}{\sqrt{\mathrm{Var}(X(t))\mathrm{Var}(X(t+\tau))}} \tag{42}$$

for an arbitrary time difference τ. As an illustration, we calculate this for the Wiener process. We know already that:

$$\mathrm{Var}(W(t)) = t \quad \text{and} \quad \mathrm{Var}(W(t+\tau)) = t+\tau.$$

We can calculate the correlation $\langle W(t)W(t+\tau)\rangle$ as follows:

$$\begin{aligned}
\langle W(t)W(t+\tau)\rangle &= \langle \int_0^t dW \int_0^{t+\tau} dW \rangle && (78)\\
&= \langle \int_0^t dW \left(\int_0^t dW + \int_t^{t+\tau} dW\right)\rangle && (79)\\
&= \langle (\int_0^t dW)^2\rangle + \langle \int_0^t dW \int_t^{t+\tau} dW\rangle && (80)\\
&= \langle W(t)^2\rangle + \langle W(t)\rangle\langle W(\tau)\rangle && (81)\\
&= t + 0 = t. && (43)
\end{aligned}$$

Thus, the correlation coefficient is:

$$C_{X(t)X(t+\tau)} = \frac{t}{\sqrt{t(t+\tau)}} = \sqrt{\frac{1}{1+\tau/t}} \tag{44}$$

Here we used the fact that the random variables $A = \int_0^t dW$ and $B = \int_t^{t+\tau} dW$ are independent, which implies their correlation $\langle AB \rangle$ is just the product of their means, $\langle A \rangle \langle B \rangle$.

From equation (44), we see that the Wiener process at time $t + \tau$ is increasingly independent of its value at an earlier time t as τ increases.

6.1 The Meaning of the Auto-Correlation Function

The function:

$$g(t, t') = \langle X(t)X(t') \rangle \tag{45}$$

is often called the two-time correlation function or the auto-correlation function ("auto" because it is the correlation of the process with itself at a later time).

If the mean of the process $X(t)$ is constant with time, and the auto-correlation function $g(t, t+\tau) = \langle X(t)X(t+\tau) \rangle$ is also independent of the time t, so that it depends only on the time difference τ, then $X(t)$ is referred to as being "wide-sense" stationary. In this case, the auto-correlation function depends only on τ, and we write:

$$g(\tau) = \langle X(t)X(t+\tau) \rangle \tag{46}$$

The auto-correlation function for a wide-sense stationary process is always symmetric, so that:

$$g(-\tau) = g(\tau). \tag{47}$$

This is easily shown by noting that:

$$g(\tau) = \langle X(t)X(t-\tau) \rangle = \langle X(t)X(t+\tau) \rangle = g(\tau). \tag{82}$$

7 Conditional Probability Density

One can always calculate the correlation $\langle X(t')X(t) \rangle$ at two times t' and $t = t' + \tau$ for some arbitrary process $X(t)$, as long as one has the joint probability density that the value of the process is x at time t and x' at time t'. Let us define the probability density as:

$$P(x, t; x', t') = P(x, t|x', t')P(x', t') \tag{48}$$

The conditional probability is the probability density for X at time t, given that X has the value x' at time t'. In fact, we already know how to calculate this, since it is the same thing we have been calculating all along in solving stochastic differential equations: the solution to an SDE for X is the probability density for X at time t, given that its initial value at $t = 0$ is x_0. To obtain the conditional probability in equation (48), all we need to do is solve the SDE for x, but this time with the initial time being t' rather than 0.

7.1 Example: Wiener Process

As an example, let us do this for the simplest stochastic equation:

$$dX = dW. \tag{83}$$

Solving the SDE means summing all the increments dW from time t' to t, with the initial condition $X(t') = x'$. The solution is:

$$X(t) = x' + \int_{t'}^{t} dW = x' + W(t - t') \tag{49}$$

This has the probability density:

$$P(x, t|x', t') = \frac{1}{\sqrt{2\pi(t - t')}} \exp\left(-\frac{(x - x')^2}{2(t - t')}\right) \tag{50}$$

To calculate the joint probability density, we now need to specify the density for X at time t'. If X started with the value 0 at time 0, then at time t' the density for $X(t')$ is just the density for the Wiener process:

$$P(x', t') = \frac{1}{\sqrt{2\pi t'}} \exp\left(-\frac{(x')^2}{2t'}\right) \tag{51}$$

Using equations (50) and (51), the joint probability density is:

$$P(x, t|x', t') = \frac{1}{\sqrt{2\pi(t - t')t'}} \exp\left(-\frac{(x - x')^2}{2(t - t')} - \frac{(x')^2}{2t'}\right) \tag{52}$$

7.2 Correlation Function

The correlation function is therefore:

$$\langle X(t')X(t)\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xx'P(x, t; x', t')dxdx' = t' \tag{53}$$

We then obtain the correlation coefficient between $X(t)$ and $X(t')$ by dividing this by the square root of the product of the variances as above.

8 Fourier Transform Formalism for Stochastic Variables

Let us define what we mean by the Fourier transform of a stochastic process. A stochastic process, $x(t)$, has many possible sample paths. Therefore, we can think of $x(t)$ as being described by a probability density over the whole collection of possible sample paths. Each one of these sample paths is a function of time, say $x_\alpha(t)$, where α labels the different possible paths. Thus, $x(t)$ is actually a random function, whose possible values are the functions $x_\alpha(t)$. Just as $x(t)$ is a random function, whose values are the sample paths, we define the Fourier transform, $x(\nu)$, as a random function whose values are the Fourier transforms of each of the sample paths. Thus, the possible values of $x(\nu)$ are the functions:

$$x_\alpha(\nu) = \int_{-\infty}^{\infty} x_\alpha(t) e^{-i2\pi\nu t}\, dt. \tag{1}$$

For a stochastic signal, $x(t)$, the total average energy in the signal is the average value of the instantaneous power, $|x(t)|^2$, integrated over all time:

$$\int_{-\infty}^{\infty} \mathbb{E}[|x(t)|^2]\, dt = \int_{-\infty}^{\infty} S(\nu)\, d\nu, \tag{2}$$

with the total energy being conserved in the ν-space also:

$$\int_{-\infty}^{\infty} |x(t)|^2\, dt = \int_{-\infty}^{\infty} |x(\nu)|^2\, d\nu. \tag{3}$$

With this definition of the Fourier transform of a stochastic process, we can now derive the proof that the autocorrelation and the energy spectrum of the stochastic process are Fourier pairs, just like we did for deterministic functions:

$$R_x(\tau) = \int_{-\infty}^{\infty} S_x(\nu) e^{i2\pi\nu\tau}\, d\nu, \tag{4}$$

where:

$$S_x(\nu) = \int_{-\infty}^{\infty} R_x(\tau) e^{-i2\pi\nu\tau}\, d\tau. \tag{5}$$

8.1 Average Power and Power Spectral Density

Let us now define the average power of a stochastic signal in the same way as for a deterministic signal, but this time we take the expectation value of the process, so as to average the power both over time and over all realizations (sample paths) of the process. Thus, the average power of a stochastic process is:

$$P = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} \mathbb{E}[|x(t)|^2]\, dt. \tag{6}$$

With the average power defined, it is useful to define the power spectral density, $S(\nu)$. This is defined as the power per unit frequency of a sample path of the process, averaged over all the sample paths:

$$S(\nu) = \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\left|\int_{-T/2}^{T/2} x(t)e^{-i2\pi\nu t}\, dt\right|^2\right]. \tag{7}$$

It turns out that the power spectral density of a wide-sense stationary stochastic process $x(t)$ is the Fourier transform of the two-time autocorrelation function. Thus:

$$S_x(\nu) = \int_{-\infty}^{\infty} R_x(\tau)e^{-i2\pi\nu\tau}\, d\tau, \tag{8}$$

where this result is called the Wiener–Khinchin theorem.

8.2 Wiener–Khinchin Theorem

The Wiener–Khinchin theorem establishes the relationship between the autocorrelation function $R_x(\tau)$ and the power spectral density $S_x(\nu)$ of a wide-sense stationary stochastic process $x(t)$. It states:

$$S_x(\nu) = \int_{-\infty}^{\infty} R_x(\tau)e^{-i2\pi\nu\tau}\, d\tau. \tag{8}$$

The inverse relationship is:

$$R_x(\tau) = \int_{-\infty}^{\infty} S_x(\nu)e^{i2\pi\nu\tau}\, d\nu. \tag{9}$$

This result shows that the power spectral density and the autocorrelation function are Fourier transform pairs.

9 Stationarity of Stochastic Processes

A stochastic process $x(t)$ is said to be stationary if its statistical properties do not change with time. Specifically:

- The mean of the process is constant:

$$\mathbb{E}[x(t)] = \mu. \tag{10}$$

- The variance is independent of time:

$$\mathrm{Var}[x(t)] = \mathbb{E}[x(t)^2] - (\mathbb{E}[x(t)])^2. \tag{11}$$

- The autocorrelation function depends only on the time difference $\tau = t_2 - t_1$:

$$R_x(t_1, t_2) = R_x(\tau). \tag{12}$$

For a wide-sense stationary process, these conditions ensure that the process exhibits time-invariant behavior in its first and second-order statistics.

10 Ergodicity in Stochastic Processes

A process is ergodic if its time averages are equal to its ensemble averages. Mathematically:

$$\lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)\, dt = \mathbb{E}[x(t)]. \tag{13}$$

10.1 Ergodicity in Mean

A process is ergodic in mean if:

$$\lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)\, dt = \mu. \tag{14}$$

10.2 Ergodicity in Autocorrelation

A process is ergodic in autocorrelation if:

$$\lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t+\tau)\, dt = R_x(\tau). \tag{15}$$

11 Properties of Gaussian Processes

A Gaussian process is a stochastic process where every finite collection of random variables has a multivariate normal distribution. This means:

$$x(t_1), x(t_2), \ldots, x(t_n) \sim \mathcal{N}(\mu, \Sigma), \tag{16}$$

where μ is the mean vector, and Σ is the covariance matrix with entries:

$$\Sigma_{ij} = R_x(t_i - t_j). \tag{17}$$

11.1 White Noise

White noise is a special case of a Gaussian process with the following properties:

- The mean is zero:
$$\mathbb{E}[x(t)] = 0. \tag{18}$$

- The autocorrelation function is a delta function:
$$R_x(\tau) = \sigma^2 \delta(\tau), \tag{19}$$
where σ^2 is the variance.

- The power spectral density is flat:
$$S_x(\nu) = \sigma^2. \tag{20}$$

12 Power Spectrum of Stochastic Processes

Different types of noise are generated by different stochastic processes. The power spectrum of a noise signal is referred to using colors.

12.1 White Noise

One type is white noise, where each component of the noise signal has a probability distribution with zero mean and finite variance and is statistically independent. This results in a noise signal with a spectral density that is even throughout all frequencies (flat power spectral density). Note that the name is drawn from white light as it contains all colors.

12.2 Red Noise (Brownian Noise)

Another type is red noise or Brownian noise, which refers to noise resulting from Brownian motion. The spectral density of this type is inversely proportional to the frequency squared. This means its power drastically decreases as its frequency increases (has more energy at low frequencies). It is called red noise because it is analogous to red light, which has a low frequency.

13 White Noise

Consider a function $f(t)$ whose integral from $-\infty$ to ∞ is finite. The more sharply peaked $f(t)$ is (i.e., the smaller the smallest time interval containing the majority of its energy), the less sharply peaked is its Fourier transform. Similarly, the broader $f(t)$ is, the narrower its Fourier transform.

Now consider a stochastic process $x(t)$. If the auto-correlation function $g(\tau) = \langle x(t)x(t+\tau)\rangle$ drops to zero very quickly as $|\tau|$ increases, then the power spectrum of the process $x(t)$ must be broad, meaning $x(t)$ contains high frequencies. This is reasonable, since if a process has high frequencies, it can vary on short time scales and, therefore, become uncorrelated with itself in a short time.

We know that the derivative of the Wiener process does not exist. However, there is a sense in which the auto-correlation of this derivative exists, which can be useful as a calculational tool. For the sake of argument, let us call the derivative of the Wiener process $\xi(t)$. Since the increments of the Wiener process in two consecutive time intervals dt are independent, $\xi(t)$ must be uncorrelated with itself whenever the time separation is greater than zero. Thus:

$$\langle \xi(t)\xi(t+\tau)\rangle = 0 \quad \text{if} \quad \tau > 0. \tag{84}$$

If we try to calculate $\langle \xi(t)\xi(t+\tau)\rangle$, we obtain:

$$\begin{aligned} g(0) &= \langle \xi(t)\xi(t+\tau)\rangle & (85)\\ &= \lim_{\Delta t\to 0} \langle \frac{\Delta W}{\Delta t}\frac{\Delta W}{\Delta t}\rangle & (86)\\ &= \lim_{\Delta t\to 0} \langle \frac{(\Delta W)^2}{(\Delta t)^2}\rangle & (87)\\ &= \lim_{\Delta t\to 0} \frac{1}{\Delta t} = \infty. & (9)\end{aligned}$$

A function with this property is the delta function $\delta(\tau)$. Let us see what happens if we assume that $\xi(t)$ is a noise source with the auto-correlation function:

$$\langle \xi(t)\xi(t+\tau)\rangle = \delta(\tau). \tag{10}$$

We now try using this assumption to solve the equation:

$$dx = \sigma dW. \tag{11}$$

The solution to (11) is:

$$x(t) = \sigma W(t). \tag{88}$$

If $\xi(t)$ exists, then we can write the stochastic equation as:

$$dW = \xi(t)dt \quad \Longrightarrow \quad dx = \sigma\xi(t)dt \quad \Longrightarrow \quad \frac{dx}{dt} = \sigma\xi(t). \tag{89}$$

Thus:

$$x(t) = \sigma \int_0^t \xi(s)ds. \tag{12}$$

13.1 Variance of $x(t)$

The variance of $x(t)$ is:

$$\mathrm{Var}(x(t)) = \langle x(t)^2 \rangle - \langle x(t) \rangle^2. \tag{90}$$

Since $\langle x(t) \rangle = 0$ (as $\xi(t)$ is a zero-mean process), then:

$$\begin{aligned} \mathrm{Var}(x(t)) &= \langle \sigma^2 \int_0^t \delta(s)ds \int_0^t \delta(v)dv \rangle & (91) \\ &= \sigma^2 \int_0^t \int_0^t \langle \delta(s-v) \rangle ds dv & (92) \\ &= \sigma^2 \int_0^t dv = \sigma^2 t. & (13) \end{aligned}$$

This is the correct answer.

13.2 Two-Time Auto-Correlation Function

The two-time auto-correlation function of $x(t)$ is:

$$\begin{aligned} \langle x(t)x(t+\tau) \rangle &= \langle \sigma^2 \int_0^t \delta(s)ds \int_0^{t+\tau} \delta(v)dv \rangle & (93) \\ &= \sigma^2 \int_0^t \int_0^{t+\tau} \langle \delta(s-v) \rangle ds dv & (94) \\ &= \sigma^2 \int_0^t dv = \sigma^2 t. & (14) \end{aligned}$$

This is also correct. Hence, we obtain the correct solution to the stochastic differential equation by assuming that:

$$\xi(t) \equiv \frac{dW(t)}{dt} \tag{95}$$

exists and has a delta auto-correlation function.

13.3 Power Spectrum of $\xi(t)$

Since the power spectrum of a process is the Fourier transform of the auto-correlation function, the spectrum of $\xi(t)$ is:

$$
\begin{aligned}
S(\nu) &= \int_{-\infty}^{\infty} \delta(t) e^{-2\pi i\nu t} dt & (96) \\
&= 1. & (15)
\end{aligned}
$$

This spectrum is constant for all frequencies ν, meaning that $\xi(t)$ contains arbitrarily high frequencies and infinitely rapid fluctuations. It also implies that the power in $\xi(t)$ is infinite. These are reflections of the fact that $\xi(t)$ is an idealization and cannot be realized by any real noise source. Because the spectrum contains equal amounts of all frequencies, $\xi(t)$ is referred to as "white" noise.

13.4 Properties of Stochastic Processes

13.5 Fokker-Plank Equation

13.6 Reverse-Time Stochastic Equation

13.7 Simple Reverse Time SDE

13.8 Numerical Methods for SDEs

14 Overview of Inverse Problems and Image Restoration

14.1 Super Resolution

14.2 Inpainting

14.3 Turbulence

14.4 Blurring

14.5 Phase Retrieval

15 Key Diffusion Model Algorithms for Inverse Problems and Image Restoration

15.1 Plug-and-Play Image Restoration

15.2 StableSR, DiffBR, BIRD, DPS, and Blind-DPS, ReSample, RePaint

Detailed analysis of specific diffusion algorithms and their application to inverse problems, with a focus on BIRD, DPS, and Blind-DPS.

15.3 Key Takeaways

Discussion of the ReSample method and other recent contributions to diffusion-based image restoration and enhancement.

16 Challenges in Medical Imaging with Diffusion Models

16.1 Addressing High Dimensionality and Stable Inference

Methods for handling high-dimensional data in diffusion models and strategies for effective image representation in medical imaging.

16.2 Cross-Modality Learning and Latent Space Optimization

Survey of cross-modality learning strategies and latent space optimization methods in diffusion models for improved adaptability and accuracy.

17 Conditional and Interpretable Image Segmentation

17.1 Overview of Controllability in Text-Image Diffusion Models

17.2 Techniques for Multi-Class Segmentations

Discussion on segmentation methods for diverse medical imaging tasks, such as Cardiac MRI, and the application of attention mechanisms.

17.3 Conditional Image Segmentation in Medical Imaging

.

Overview of techniques for conditional segmentation, including multi-class segmentation for medical imaging applications.

18 Knowledge Distillation in Diffusion Models

18.1 Plug-and-Play Diffusion Distillation

18.2 Adverserial Diffusion Distillation

18.3 CoDi: Conditional Diffusion Distillation

19 Adapting Diffusion Models for Various Domains

19.1 Discrete, Invariant, Manifold Structural Data

.

19.2 Low-Rank and Sparse Structure

Techniques for incorporating low-rank and sparse structures in diffusion models for signal processing, highlighting control strategies and model modifications.

20 Controllable and Semantically Consistent Generation

20.1 Techniques for Semantic Consistency

Strategies to achieve semantic consistency in generated images, including attention maps and embeddings for meaningful generation.

20.2 Semantic Correspondence Applications

Applications of semantically consistent generation, with a focus on unsupervised learning and semantic correspondence tasks.

21 Conclusion

21.1 Summary of Key Findings

A summary of the key insights from the survey and their implications for diffusion models in medical imaging and signal processing.

21.2 Limitations and Open Challenges

Discussion of current limitations and potential challenges for future research in these domains.

A Appendices

A.1 Code Resources

Links to relevant GitHub repositories, tools, or resources.

A.2 Suplementary Material

Supplemental proofs and derivations supporting the discussed methodologies.

A.2.1 Tweedie's Formula

A.2.2 Total Variation Regularization

A.3 Additional Figures

Figures and diagrams for extended illustration of key concepts.

Acknowledgments

This work has been supported and guided by Dr. Hassan Mohy-ud-Din and Dr. Muhammad Tahir. Their mentorship and expertise have been instrumental in developing this research.

References